

Eötvös Loránd Tudományegyetem

Bölcsészettudományi Kar

Doktori disszertáció tézisei angol nyelven

Bernáth László

Morális felelősség, érdem és kontroll –

A morális felelősség metafizikája

Filozófiatudományi Doktori Iskola

Dr. Boros Gábor, DSc, Filozófiatudományi Doktori Iskola vezetője

Analitikus Filozófiai Program

Dr. habil. Zvolenszky Zsófia, PhD, Analitikus Filozófiai Program vezetője

A bizottság tagjai és tudományos fokozatuk:

Dr. E. Szabó László, DSc

Dr. Huoranszki Ferenc, DSc

Dr. Réz Anna, PhD

Dr. habil. Szalai Judit, PhD

Dr. habil. Lautner Péter, PhD

Dr. Márton Miklós, PhD

Dr. habil. Schmal Dániel, PhD

Témavezető:

Dr. habil. Tózsér János, PhD

Budapest, 2016

PhD Theses

Moral Responsibility, Desert and Control – The Metaphysics of Moral Responsibility

Doctoral Dissertation

Bernáth László

2016

My dissertation has two main aims. First, I argue that moral responsibility is not compatible with determinism. Second, I attempt to describe the outline of the metaphysics of moral responsibility. I claim that there are morally responsible people only if agents are substances who are irreducible to the sum of their properties. Furthermore, I argue, there is moral responsibility only if agents have decisions that are made in situations in which natural laws and the properties of the agents outline only the alternatives while the agents as substances determine which option takes place. These free decisions have a curious feature: the open alternatives have no metaphysically robust ground-floor probabilities prior to the decision.

In the first part of the first chapter, I attempt to confute the main variants of normative compatibilism using ethical examples and thought experiments (1.2.). I call theories normative compatibilism according to which determinism and moral responsibility are compatible with each other because one is blameworthy in any situation in which blaming the agent is useful (1.2.1.), appropriate in a contractualist framework (1.2.2.), or fair (1.2.3.).

In the second part of the first chapter after the rejection of these theories, I analyze the concept of desert showing the main features of the desert-theory of moral responsibility. According to this approach, one is morally responsible for X if and only if she *deserves* moral blame or praise for X insofar as X was morally bad or good (1.3.). Besides the fact that the notion of desert has received less attention than it would be reasonable regarding the structure of the problem of moral responsibility, I consider this analysis of the notion of desert (1.3.1-1.3.2.) to be especially relevant since the desert-theory of moral responsibility provides the ethical basis of incompatibilism.

In section 1.4, after I describe the main features and different versions of the desert-theory of moral responsibility (1.3.3.), I face the strongest argument against it. According to the argument, desert-theory of moral responsibility has an unacceptable consequence. Namely, blaming or punishing the offender is intrinsically good because she deserves

suffering. Firstly, I show the real consequence of the approach. Secondly, I argue this consequence is not unacceptable at all.

In the end of the chapter, I provide two further arguments for the desert-theory of moral responsibility which are not based on failures of other approaches. The first argues that the desert-theory fits the characteristics of moral blame and praise well (1.5.1.). The second claims that one has a plausible explanation of some features of the discourses about moral responsibility only if she chooses the desert-theory of moral responsibility (1.5.2.).

In the second chapter, I refute the compatibilist theories which can be interpreted in such a way that they claim that having specific properties can be the ultimate desert base of blame and praise even if it is not the agent's fault that she has the property in question. I claim that Strawsonian and one-way compatibilism should be interpreted in this way. The former claims the following: *because* the actual moral practice justifies itself, agents deserve blame for having a morally problematic property even if it is not their fault. The reason is simple: it is compatible with the everyday moral practice that people blame others for having a property regardless of what is the cause of having the property in question. According to the latter theories, the reason why moral responsibility is compatible with determinism is that agents are morally responsible for having some morally relevant properties and for actions which stand in the appropriate relation with the mentioned properties even if these properties, for example having a particular bad characteristic feature, are not the agents' fault.

My main counter-argument against the naturalist accounts of moral responsibility is that moral relativism follows from this kind of theory and moral relativism is implausible with regard to moral responsibility (2.2.). If one claims that the fundamental parts of our moral practices justify themselves, and she is aware of that blaming practices of different cultures are in conflict, she has to accept that every moral practice with regard to moral responsibility is right. So in different cultures, people deserve blame for different actions, which means a person's blameworthiness depends on the culture in which the person lives. It seems to be implausible.

After criticizing the naturalist accounts of moral responsibility, I reconstruct the one-way compatibilist theories of moral responsibility (2.3.). In the following section (2.4.), I argue that although these theories are seemingly able to explain well how agents can be responsible for many different types of actions and properties, such as omissions, desires, beliefs, emotions, characteristics and so on, this advantage is illusory since the provided explanation does not fit our fundamental ethical convictions and practices (2.4.).

In section 2.5., I construct two arguments against one-way compatibilism. The first one is the isolation-argument which has the aim to show how implausible the claim is that agents may be responsible for their characteristics and for their actions which are inevitable consequences of their characteristics regardless of the antecedent events which explain why the agents have the morally problematic characteristic (2.5.2.). The second argument is based on the analysis of moral blame (2.5.3.). According to the argument, moral blame includes the thought that the basis of blame is the blamed person's fault. That is why moral blame can be deserved only if the basis of blame is in fact the blamed person's fault. Thus, the one-way compatibilism is wrong because it follows from the theory that one can be blameworthy for properties which are not her fault. This chapter argues against two notable versions of compatibilism; besides, it is relevant with regard to the whole train of thought of the dissertation because it makes clear by the analysis of moral blame why the incompatibilists are right about that agents must have strong control over the basis of blame to be blameworthy. If agents are blameworthy only for their faults, and only features which are strongly controlled by the agent can be the blamed person's fault, any agent can be blameworthy only for the features which are strongly controlled by the agent (2.6.).

In the third chapter, I directly argue for the two main claims of the dissertation. The first one: if one holds a plausible interpretation of natural laws, she has to claim that deterministic laws are incompatible with moral responsibility. The second one: enhanced control which is needed for moral responsibility is possible only if agents are such substances whose causal powers cannot be reduced to the causal powers of their properties, and they are able to control the usage of their specific causal power in a way that it is not possible to attribute metaphysically robust ground-floor probabilities to any open alternatives.

I start from a widely accepted thought according to which anything can be fault of an agent only if the agent exercises strong control over their actions, omissions or properties (3.1.). That is the reason why the understanding of control and enhanced control are the key for the problem of moral responsibility. Accordingly, in the next section (3.2.), I analyze the notion of control. I argue for two, closely related claims. The first one says that the best control-theory is the effect-theory of control (3.2.1–3.2.5.). According to this approach, the control-relation is based on to what extent the controller is able to affect the controlled being. The second claim is that only one explanation can satisfy the fact why humans, in contrast to – for example – present-day machines, have enhanced control, which is the condition of moral responsibility. Namely, the explanation according to which only humans are able to control which optional course of events is started by them (3.2.6-3.2.7.) is that one. The

control that is the *ultimate and direct* control over which course of events is started by the agent is the crucial condition of having enhanced control. After the analysis of the notion of control, I explain to what extent ultimate and direct control is the condition of enhanced control and moral responsibility (3.3.).

Based on the results of the analysis of control, I outline the metaphysics of free decisions which are the ultimate source of moral responsibility (3.4.). I argue for that one can comprehend what the difference is between free decisions and random events if it is supposed that the agent is a substance who has more than one open alternative with regard to how she exercises her special irreducible causal power to determine which course of events starts. After describing the main features of the theory (3.4.1.), I attempt to show that two claims about the metaphysics of decisions are true. Firstly, the presence of ground-floor probabilities and the ultimate direct control are incompatible (3.4.2.). Secondly, the lack of ground-floor probabilities can be explained by the metaphysics of agent-causation only. It means that neither the solution of non-causal libertarianism nor the account of event-causal libertarianism work. In the first case, if the non-causal libertarian claims that free decisions have no causes, she cannot explain plausibly what the difference is between random events without causes and free decisions. In the latter case, if the event-causal libertarian holds that free decisions are indeterministically caused by events, she is unable to explain what the relevant difference is between indeterministic chain of events and free decisions.

After describing the metaphysics of free decisions, I return to dealing with the compatibility problem of moral responsibility and determinism (3.5.). Based on the result of the previous analysis claiming that ultimate direct control is a condition for moral responsibility, I argue neither two-way compatibilism (3.5.1.) nor semi-compatibilism (3.5.2.) are able to explain moral responsibility. Although two-way compatibilism accepts the thesis that open alternative possibilities are necessary conditions for moral responsibility, it cannot show how the *ultimate direct* control and determinism could be compatible with each other. If, as I argue, possessing ultimate direct control means that the agent is able to start one of the open alternatives of course of events whereas the past, the natural laws and the given situation are fixed, the deterministic natural laws rule out this kind of control because they allow to start only one course of events without breaking the natural laws in every situation.

Even though semi-compatibilism does not face this problem since it holds neither free will nor ultimate direct control is a necessary condition of moral responsibility, it is unable to show how agents have such enhanced control which can explain why only humans, as opposed to – for example – machines, are able to be responsible for properties, actions and

omissions. Although semi-compatibilism is right about that agents' reason-responsiveness means a relevant difference between present-day machines and agents, this difference is able to provide explanation why people can control more types of events than machines.

In the last section, I summarize the train of thought which is against compatibilism and can be traced through the whole dissertation (3.5.3.). According to this argument, determinism rules out moral responsibility because it allows only "mechanistic" reactions. It means that if the natural laws are deterministic, every being can react to any stimulus in only one way in any given situation. Consequently, every being controls to the same extent which course of events is facilitated as a reaction to a stimulus in the given situation by the being in question. And if every being controls their reactions to the same extent, humans do not have a stronger control over their reactions than our favorite tools which are considered that reliable just because we know which one is the only possible way for them to react to any stimulus. However, if we are unable to control our reactions better than our tools are able to control their reactions, we are not responsible for our reactions as well as our reliable tools are not morally responsible for anything.